

# Analyzing transit service quality evolution using decision trees and gender segmentation

R. de Oña<sup>1</sup> & J. de Oña<sup>2</sup>

<sup>1</sup>*TRYSE Research Group, Department of Civil Engineering,  
University of Granada, Spain*

<sup>2</sup>*TRYSE Research Group, UGR, Spain*

## Abstract

The analysis of service quality in public transportation has become very popular in the last few years, when transport planners and managers have begun to focus all their efforts on a continuous improvement of the service quality. Knowing the passengers' perceptions about different characteristics describing the service allows transport managers to decide where they have to apply their actions. However, passengers' perceptions and requirements about public transport services are very heterogeneous among users, and also over the time. Segmentation is normally carried out in order to reduce this heterogeneity, and with an annually or six-monthly frequency, detailed analysis are performed in order to identify the most significant changes in passengers' opinions. In this research work, a segmentation based on the gender of passengers is developed, in order to discover and compare the key factors affecting the perceptions of men and women about the quality of service. Data from three consecutive Customer Satisfaction Surveys (2009–2011) developed in the metropolitan transit service of Granada (Spain) are used, and the evolution of the most influence factors for men and women across the years is shown. For this purpose, a novel methodology for the analysis of service quality is presented. Decision Trees, and particularly the CART algorithm, has been applied. This is a non-parametric technique, with no pre-defined underlying relationship between the target (dependent) variable and the predictors (independent variables), which has several advantages over parametric models. The main results shown that for men the most influence variable changes among Safety, Punctuality and Timetable, while for women, the key factor varies among Speed, Frequency, and Timetable. *Keywords: service quality, decision trees, CART, heterogeneity, gender segmentation.*



## 1 Introduction

Delivering high quality services becomes an essential requirement for public transport providers, given that an increase in customer satisfaction translates into retained markets, increased use of the system, newly attracted customers, and a more positive public image [1]. In fact, when a customer is satisfied about one company's product or service may introduce 1–3 other customers, otherwise if he has a negative assessment, he may tell 11–13 relatives or friends [2].

In the context of public transportation, Service Quality (SQ) is usually measured by the point of view of users. Numerous researchers have highlighted the importance of passengers' perceptions, since they consider that passengers are the real judges of the service. Then, users' perceptions about a series of attributes describing the performance of the service provide managers and operators a reliable measure about the quality of the service supplied. However, passengers' opinions and requirements about these attributes are very heterogeneous among users. This heterogeneity depends on a large variety of factors, such as: the qualitative nature of certain aspects that characterize the services, the attitudes passengers have towards the use of public transportation, the different ways of viewing aspects of the service, and the social and economic characteristics of passengers and their preferences [3].

Stratifying the sample into more uniform segments can be a good solution for reducing this heterogeneity and identifying which are the real key factors influencing passengers' overall perceptions about service quality. Usually this segmentation is based on social and demographic characteristics of the passengers (i.e. age, gender, use frequency and so on), and subsequently, specific models are generated with these more homogeneous segments [4].

However, the most important factors for passengers also change across the time, mainly due to the actions service providers make in the system (such as changes of timetables, changes on the route of a line, frequencies, acquisition of new vehicles, new fares, etc), or maybe due to punctual incidents that are out of control of service provider (accidents, traffic jams, modifications in the route for constructions, etc). Identifying the most significant changes in passengers' opinions provides service operators with a deeper understanding about the consequences of previous actions or incidents in the system.

The objective of this research work is to identify the evolution over the years in the main attributes influencing passengers' overall SQ perceptions. Data from three consecutive Customer Satisfaction Surveys (2009–2011) delivered for the bus transit service of the metropolitan area of Granada were analyzed, and in order to reduce the heterogeneity present in their opinions, a gender segmentation was developed. A decision tree approach is used, and particularly, the Classification and Regression Tree (CART) algorithm was applied. Decision trees methodology is a non-parametric technique with no pre-defined underlying relationship between the dependent variable and the independent variables. It has been widely employed in different fields of research (business administration, agriculture, industry, and engineering). However, it is a novel methodology for analyzing SQ, and particularly, SQ in public transportation. Only de Oña *et*

*al.* [5] applied this methodology for identifying the key factors affecting a bus transit SQ, and de Oña *et al.* [6] evaluated the quality of a rail service in the North of Italy. Both research works obtained successful results.

After this brief introduction, the data for the experimental context and a description of CART methodology is presented in the second section. The third section is about the results obtained from the application of the CART methodology, and finally we propose the conclusions.

## 2 Materials and methods

### 2.1 Data

The data used in this study come from various Customer Satisfaction Surveys (CSS) conducted by the Transport Consortium of Granada in the metropolitan transit service. Since the year 2006, the Transport Consortium has developed an annually survey for knowing how satisfied are passengers with the bus transit service provided in the metropolitan area of Granada (Spain).

The CSSs gathered information of different nature, such as socio-economic characteristics (sex, age, occupation), travel behavior (travel reason, use frequency, type of ticket, complementary modes from origin to train station and from train station to destination, private vehicle available), perceived SQ attributes (specifying the perceived quality of 13 attributes describing the service, indicated in Table 1) and a global evaluation of SQ. An eleven-point numeric scale (from 0 to 10) was used for measuring the overall evaluation of SQ, and the perceived SQ attributes. To find out more applicable decision rules, the target variable (overall SQ evaluation), was re-coded in a reduced semantic scale. It was a three semantic scale comprising the rates from 0 to 4 as “Poor”, from 5 to 7 as “Fair”, and from 8 to 10 as “Good”.

In this paper, data from the surveys carried out in the years 2009-2011 were analyzed. The number of valid records for this period was 966, 613 and 1603 for the years 2009–2011 respectively. Table 1 shows the average satisfaction rates calculated from the values expressed by the users about the 13 service attributes, for each year under study and distinguishing among men and women.

According to the average satisfaction rates suggest that people are not very dissatisfied with the service, having all the attributes an average rate higher than 5,5, over the years and across the gender segmentation. On the whole, the service attributes considered as the most satisfying regard Courtesy, Safety, Cleanliness, Temperature and Accessibility (with values higher than 7), while the attributes judged as the less satisfying concern Fare, Frequency and Timetable. By observing the satisfaction rates among men and women, the main difference found is that in general, men are more satisfied with almost all the attributes than women. Regarding the evolution of the satisfaction rates across the years, men had higher satisfaction rate in 2010, whereas in 2011 their perceptions had decreased. In the case of women, it can be observed that there is a light improvement in their perceptions each consecutive year.



Table 1: Average satisfaction rates.

<i>Service Quality Attributes</i>	<i>Satisfaction Rates</i>					
	<i>Men</i>			<i>Women</i>		
	2009	2010	2011	2009	2010	2011
Information	6.16	6.87	6.65	6.18	6.19	6.79
Punctuality	6.74	7.86	7.70	6.69	7.68	7.76
Safety	7.29	7.88	7.67	7.25	7.49	7.73
Courtesy	7.88	8.12	7.95	7.79	7.99	7.98
Cleanliness	7.13	7.71	7.68	7.07	7.65	7.63
Space	6.87	7.49	7.44	6.86	7.18	7.37
Temperature	7.07	7.71	7.68	7.12	7.67	7.69
Accessibility	7.11	7.92	7.35	7.03	7.39	7.41
Fare	5.94	6.67	6.34	5.82	5.99	6.40
Speed	6.75	7.44	7.11	6.56	7.07	7.26
Frequency	6.09	6.46	6.57	5.85	5.85	6.41
Proximity	6.95	7.40	7.07	6.80	7.03	7.30
Timetable	6.46	6.65	6.58	6.24	5.89	6.38

Then, Decision Trees were built using the information of the perceived SQ attributes as independent variables and the overall SQ evaluation as the dependent variable. Two different models were generated for each year under study, according to the gender segmentation, for identifying changes in the main factors affecting passengers' opinion about SQ. Finally, six different models were generated.

## 2.2 Methodology

Decision trees are an important data mining technique used for the classification and prediction of a class variable. This technique has the ability of discover useful patterns in data that allows us to make predictions on new data. According to the nature of the target variable, a classification tree or a regression tree can be developed (for discrete and continuous variables respectively). Both types of trees can be modeled using the CART method. In this study, the target variable is discrete (Overall SQ Evaluation), and thus, a classification tree is used.

A decision tree is represented by nodes and branches. The building-process starts selecting an attribute to place at the root node and create two or more branches for the different values of the attribute, splitting up the full data set into smaller subsets. Then the process can be repeated recursively for each branch, using only those cases that actually reach the branch. Each subset is split recursively until all of them are pure (when the cases in each subset are all of the same class) or a stopping criteria has been satisfied. Then the maximal tree that overfits the data is created. To decrease its complexity, the tree can be pruned using different pruning algorithms.

CART methodology creates binary trees using the Gini reduction as the splitting criteria and a cost-complexity algorithm for pruning the tree. The Gini reduction used for splitting the tree measures the "worth" of each split in terms

of its contribution toward maximizing the homogeneity through the resulting split. If a split results in splitting of one parent node into B branches, the “worth” of that split may be measured as follows:

$$\text{Worth} = \text{Impurity (Parent node)} - \sum_{b=1}^B P(b) * \text{Impurity (b)} \quad (1)$$

where Impurity (Parent node) denotes the Gini measure for the impurity (i.e., non-homogeneity) of the parent node, and  $P(b)$  denotes the proportion of observations in the node assigned to a branch b. The impurity measure, Impurity (node), may be defined as follows, in which I is the number of classes in the target variable:

$$\text{Impurity (node)} = 1 - \sum_{i=1}^I \left( \frac{\text{number of class } i \text{ cases}}{\text{all cases in the node}} \right)^2 \quad (2)$$

When the maximal tree is created, it is pruned using the cost-complexity algorithm, which is based on removing the branches that add little to the predictive value of the tree. After pruning a branch, if the increase in the misclassification cost is sufficiently lower than the decrease in the complexity cost, the branch will be pruned, and a new tree is created. Among the pruned trees, the optimal tree is selected. It is the one that has the least misclassification cost on the testing dataset. More detailed description of CART analysis and its applications can be found in Breiman *et al.* [7].

One of the most valuable outcomes provided by CART analysis is the value of the standardized importance of independent variables, which reflects the impact of such predictor variables on the model.

### 3 Results

Six different decision trees were built to classify the overall SQ of the bus transit service of the metropolitan area of Granada. Through the six models generated, the attributes that play a key role in the classification of the target variable across the passengers' gender and across the years were identified.

The CART building-process used the 80% of the sample for training the model and the rest (20%) as the testing sample. The accuracy rate of the generated models among the different years (2009–2011) was: for MEN, 79.2%, 79.4% and 83.0% respectively, and for WOMEN, 67.7%, 67.4% and 74.9%. These values are acceptable and higher than the values obtained in other studies in which decision trees were applied with similar objectives [5, 8].

The decision trees built for the year 2009 are shown in the Figure 1 and the Figure 2 for explaining their usefulness. The tree developed was considerably different for men and for women, as well as the attributes identified as the most important. The variable that split the root node as well as the number of levels, number of nodes and terminal nodes generated were very different among the two trees. While the variable that split the root node for men was Punctuality, for women this variable was the Timetable.

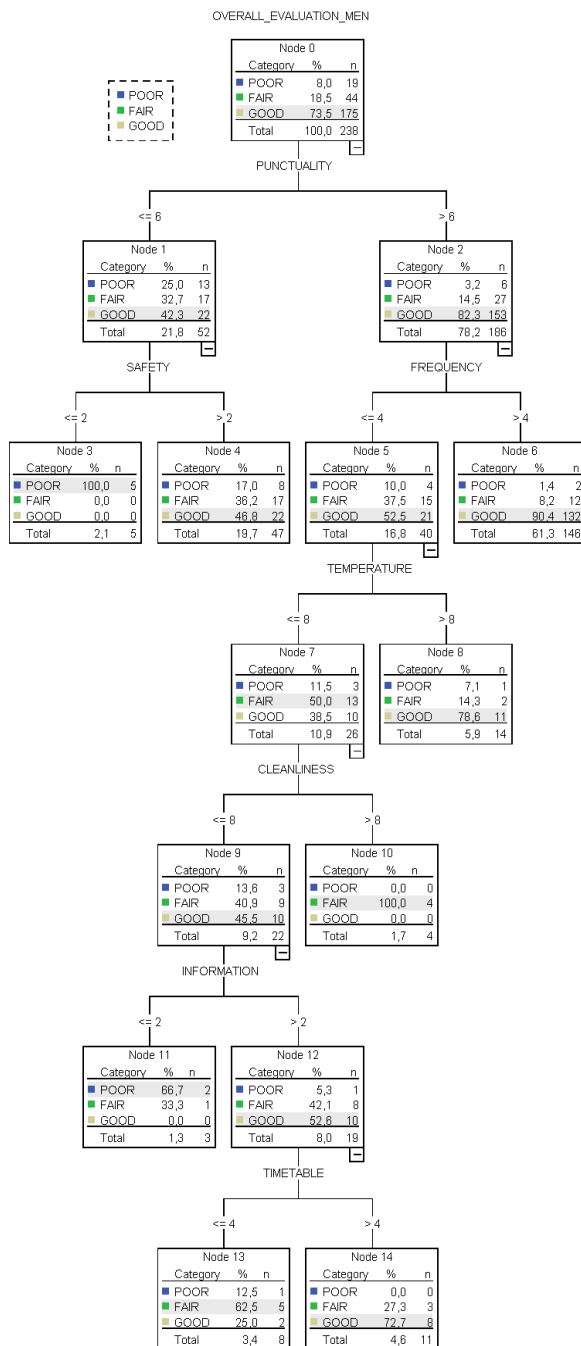


Figure 1: CART generated for men for the data of 2009.

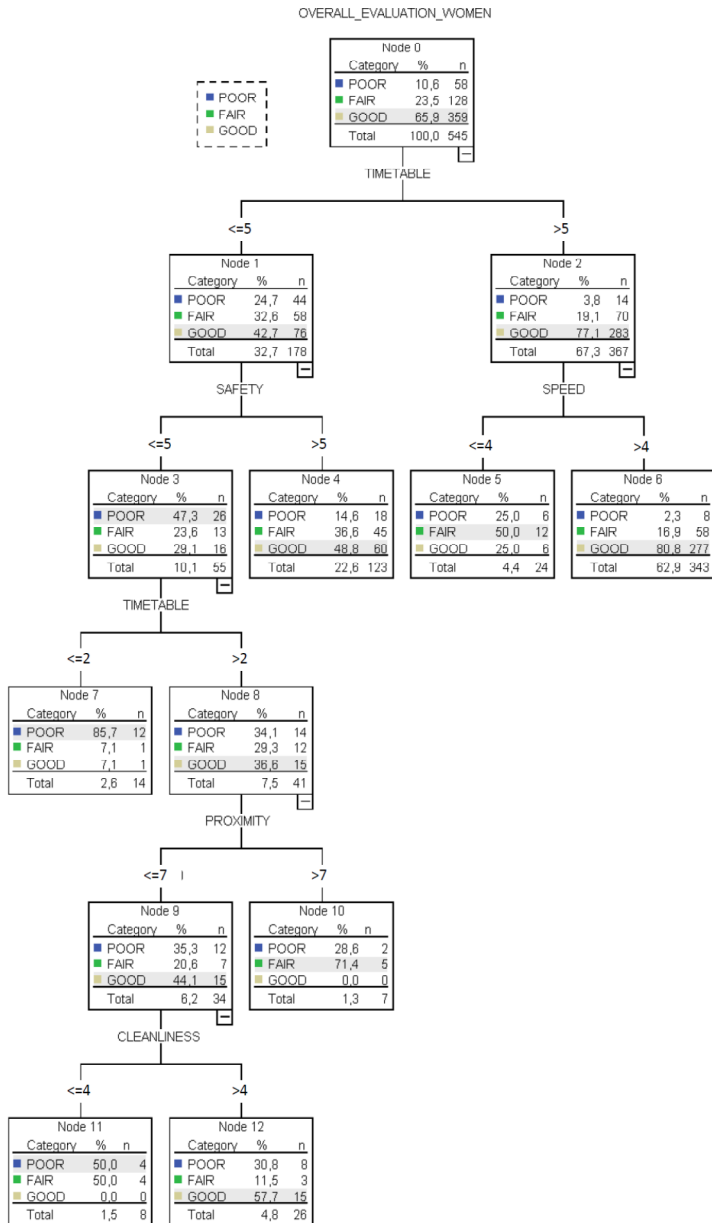


Figure 2: CART generated for women for the data of 2009.

The tree generated for the men (see Figure 1), produced six levels (depth below the root node), 15 nodes and 8 terminal nodes or leaves. The terminal nodes predicted one class of the target variable, from which, two predicted

“Poor” (Nodes 3 and 11), two “Fair” (Nodes 10 and 13) and the rest four terminal nodes (Nodes 4, 6, 8 and 14) a “Good” Overall Evaluation. Punctuality, Safety, Frequency, Temperature, Cleanliness, Information and Timetable were the attributes used as splitters, and the best thresholds for splitting them were found. Following the path from the root node to a terminal node, useful “If-Then” decision rules are found. For example, the Node 3 predicts a Poor Overall Evaluation of the service, with a probability of the 100%, when various conditions are complied. “*IF Punctuality is perceived by passengers with a value minor than acceptable ( $\leq 6$ ), and the Safety is considered as awful ( $\leq 2$ ), THEN, the overall evaluation about the service quality will be Poor*”. On the contrary, the Node 8 predicts a Good Overall Evaluation with a probability of the 78.6%, when “*IF Punctuality is perceived as good ( $> 6$ ), and Frequency is considered minor than sufficient ( $\leq 4$ ), but Temperature is perceived as performance very good ( $> 8$ ), THEN, the overall evaluation will be Good*”.

In the Figure 2 is displayed the tree generated for women. It produced 5 levels, 13 nodes and 7 terminal nodes. Nodes 7 and 11 predicted a Poor Overall Evaluation, Nodes 5 and 10 a Fair evaluation and Nodes 4, 6 and 12 classified the target variable as Good. The variables used as splitters also differ from the ones used in the tree for men. They were Timetable, Safety, Speed, Proximity and Cleanliness. It can be highlighted that in this model, Proximity takes part of the tree as one of the conditions for the decision rules. It could be due to the fact that women are the most frequent users that transport a trolley for doing the shopping or for carrying their children, then, they maybe are more worried about the distance they have to move to achieve the bus stop. In this tree, also useful decision rules are discovered. For example, with a probability of 85.7% (Node 7), the Overall Evaluation will be predicted as Poor when “*IF Timetable is perceived as awful ( $\leq 2$ ), and Safety also is considered minor than sufficient ( $\leq 5$ ), THEN, the Overall Evaluation will be Poor*”.

The trees generated for the following years (2010 and 2011) also were built with different structures and splitters. The tree created for men with the data of 2010, used as splitters the Safety, Punctuality, Information, Frequency and Fare. It achieved good probabilities in the predictions on its six terminal nodes, with values over 64.7% in all them. On the contrary, the tree developed for women in the same year, had a large size, consisting of several nodes and terminal nodes (21 and 11 respectively). In this tree, the attribute Space is used as splitter while it was not considered in the tree for men. In the year 2011, the trees built for men and women were more homogeneous, with sizes very similar (11 and 13 nodes respectively), and using almost the same variables as splitters. In the case of the tree generated for women, it used, in addition to the splitters used for men, the Accessibility for increasing the purity of the child nodes.

One of the most important results obtained by these models is the importance of the variables. This is achieved by using the importance index [9], of which a standardized form has been used in this paper. Table 2 shows the standardized importance rates deduced by the six models built. Only the first five attributes have been displayed in order to simplify the outcomes.



Table 2: Standardized importance rates.

<i>Men</i>			<i>Women</i>		
2009	Safety	100,0%	2009	Speed	100,0%
	Punctuality	75,1%		Timetable	75,2%
	Frequency	55,5%		Safety	73,2%
	Temperature	47,4%		Frequency	67,2%
	Fare	42,4%		Space	57,2%
2010	Punctuality	100,0%	2010	Frequency	100,0%
	Safety	65,1%		Fare	96,0%
	Information	64,0%		Information	91,0%
	Space	37,1%		Accessibility	82,2%
	Frequency	28,6%		Timetable	71,6%
2011	Timetable	100,0%	2011	Timetable	100,0%
	Temperature	79,3%		Temperature	82,4%
	Cleanliness	52,0%		Cleanliness	72,8%
	Information	49,4%		Safety	53,1%
	Safety	35,4%		Accessibility	45,4%

By analyzing these results, large differences can be noticed. The most important variable identified by the models for men, changes every year. In 2009 the attribute Safety is the most important factor for the model, in 2010 is the Punctuality and in 2011 the Timetable. However, these are not the unique variables that have a high influence on the model. If we observe the attributes that have standardized importance values that exceed 50%, also other variables should be taken into account. For the years 2009 and 2010, Safety and Punctuality are in the highest positions, but also Frequency and Information should be taken into account in these years respectively. For the year 2011, Temperature and Cleanliness take also an important role in the model.

On the other hand, the most important factors deduced by the models built for women, also changes across the years. In 2009, Speed, Timetable, Safety, Frequency and Space are the most influence attributes for the model. All of them have values over 50%. In 2010, Frequency and Timetable are also in the highest positions; however Fare, Information and Accessibility also become crucial in this year. In 2011, Timetable, Temperature and Cleanliness are the most important variables as happen with men. This could indicate that maybe in 2011, service operators should pay more attention to this three attributes, because in previous years they were not very important, but in 2011, they became important not only for men, but also for women.

It should be pointed out that for men, few variables are important. Only three have achieved a standardized importance value higher than 50% each year. In the case of the women, the number of influence variables is higher, and also different, demonstrating that the way of creating the perceptions about the SQ is different between men and women. Then, in order to reduce the heterogeneity present in passengers' opinions, they should be analyzed separately.

## 4 Conclusions

This study has demonstrated that Decision Trees is a suitable methodology for predicting Service Quality in Public Transportation. Particularly, in this paper, CART algorithm has been applied for analyzing the changes produced through a set of years (2009-2011), on the key factors affecting men and women when they evaluate the bus service quality of the metropolitan area of Granada.

Six different models were generated, and high precision rates were achieved (values over 67.4%, which are higher than in other CART studies with similar objectives). In this research 13 attributes describing the bus transit service were used and the variables that played the most important role each year and through the different groups of the segmentation (men and women), were easily identified. For men, the most important attribute changes from Safety, Punctuality and Timetable, while for women, it changes among Speed, Frequency and Timetable. Moreover, the number of attributes influencing the overall service quality is higher for women than for men.

Then, in this research, we have verified that the main attributes affecting service quality tend to change, depending on the segment under study, and also on the period of time analyzed. For this reason, it demonstrates that segmentation and evolution analysis, are important issues for a better understanding of passengers' perceptions and requirements, leading to a higher homogeneity of their opinions, and as a result, better predictions in the models.

Many studies have analyzed bus service quality using several approaches, however, CART methodology, which is a non-parametric model with a novel used for predicting the overall service quality, has a number of benefits compare to the other widely used parametric models. The main advantage is that CART does not need to establish a functional relationship between variables as ordinary statistical modeling techniques. In regression analysis, if the model is misspecified, the estimated relationship between dependent variable and independent variables as well as model predictions will be erroneous. Another important advantage is that the outcomes of the analysis are represented in tree graphics, which make the results easy to understand. Following the paths of the tree model, useful decision rules can be extracted, and valuable information is provided. Attending to this information, public transport managers and operators could determine the priority of the measures that are going to develop in the service. In addition, CART methodology allows using many explanatory variables, and the weight of them in the model is determined.

## Acknowledgements

This study is sponsored by the Conserjería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) through the Excellence Research Project denominated "Q-METROBUS-Quality of service indicator for METROpolitan public BUS transport services". The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.



## References

- [1] Transportation Research Board (1999) A Handbook for Measuring Customer Satisfaction and Service Quality, TCRP Report 47.
- [2] Johns, M. 1995. Customer satisfaction: the case for measurement, *Journal of Audiovisual Media in Medicine* 18(1): 17–32.
- [3] Eboli, L., & Mazzulla, G. (2011). ‘A methodology for evaluating transit service quality based on subjective and objective measures from the passenger’s point of view’. *Transport Policy*, 18, pp. 172–181.
- [4] Dell’Olio, L., Ibeas, A., and Cecín, P. (2010). ‘Modelling user perception of bus transit quality’. *Transport Policy*, 17(6), pp. 388–397.
- [5] de Oña, J., de Oña, R., and Calvo, F.J. (2012). A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications*, 39, pp. 11164–11171. <http://dx.doi.org/10.1016/j.eswa.2012.03.037>
- [6] de Oña R, Eboli L, Mazzulla G. 2012. Key factors affecting rail service quality. a decision tree approach. XIX Conference SIDT, Padua, 18-19 October, 2012.
- [7] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1998). *Classification and regression trees*. Chapman & Hall/CRC.
- [8] Wong J. and Chung P. (2007). Managing valuable Taiwanese airline passengers using knowledge discovery in database techniques, *Journal of Air Transport Management*, 13 (6), pp. 362–370.
- [9] Kashani AT, Mohaymany AS. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Sci* 2011; 49:1314-1320.

